



Les biais lors de la collecte des données

Méthode
d'échantillonnage

Prise en compte de
l'environnement

L'enquêteur

L'enquêté

Jeudi 04 avril 2019



Chargées de Mission Qualité : AUBERT Magali : aubert@supagro.inra.fr ; MORIN Pascale : pascale.morin@cirad.fr





Biais d'échantillonnage

1. Echantillon non probabiliste
 1. Commodité / à l'aveugle
 2. Volontaire
 3. Par quota
 4. Effet boule de neige
2. Echantillon probabiliste
 1. Aléatoire
 2. Systématique
 3. Par degré
 4. Probabilité proportionnelle à la taille
 5. Stratifié
 6. Grappe
3. Ce qu'il vous faut retenir



Quelques éléments de rappel

- ❑ Population : toutes les personnes auxquelles on s'intéresse
- ❑ Echantillon : un sous-groupe de cette population
- ❑ Echantillon représentatif : les individus de mon échantillon représentent la population le plus fidèlement possible du point de vue de ses caractéristiques
- ❑ Pourquoi échantillonner ?
 - ❑ Enquêter la population est trop coûteuse (temps, coûts...)
 - ❑ Le chercheur ne dispose que de peu de temps
 - ❑ La population est difficilement accessible



Quelle que soit la méthode d'échantillonnage suivie ...

La définition / la taille de l'échantillon doit permettre de :

- Fournir des informations et des résultats cohérents avec ceux qui seraient obtenus si l'ensemble de la population avait été enquêtée.
- Permettre une lecture représentative du phénomène étudié



Erreur d'échantillonnage

Erreur = écart entre :

- Les résultats obtenus avec l'échantillon
- Les résultats que nous aurions eu en interrogeant l'ensemble de la population

Plus la taille de l'échantillon est grande et plus l'erreur d'échantillonnage diminue



Pour comprendre les méthodes d'échantillonnage :
1 exemple

Enquête de satisfaction auprès des voyageurs de la SNCF



+ Echantillon non probabiliste



Echantillon par commodité / à l'aveugle

Exemple : Par convenance, l'enquêteur interroge les personnes présentes dans certaines gares de France

On applique aucune méthode d'enquête particulière.

Méthode facile à utiliser et représentative uniquement si la population est homogène.



Biais en cas de population non homogène



Echantillon volontaire

Exemple : Je demande aux usagers de la SNCF s'ils sont intéressés et donc volontaires pour participer à mon enquête

Auto-sélection des individus qui composent l'échantillon

Les individus manifestent un intérêt pour une question de recherche

Facile à mettre en œuvre



Seules les personnes intéressées se manifestent et elles ont potentiellement des caractéristiques différentes de l'ensemble de la population



Echantillon par quota

Exemple : Je me mets dans chaque gare et je prends les 10 premiers voyageurs qui entrent dans le train

On définit des groupes sur la base de critères.

On interroge les individus au sein de chaque groupe selon un critère à priori « neutre »

A priori ressemble à une stratification dont la population enquêtée pourrait être considérée comme représentative de la population totale



**Le critère de sélection des individus au sein des groupes n'est pas aléatoire
– pas de représentativité de l'échantillon**



Echantillon par effet boule de neige

Exemple : Je sélectionne les porteurs de carte SNCF et leur demander d'interroger les personnes qu'ils connaissent et prennent aussi le train.

On identifie un individu de départ et par incrémentation on compose notre échantillon

Chaque individu (i) va permettre la sélection d'autres individus (j)

Possible quand on ne connaît pas la population



Le choix du 1^{er} individu conditionne la composition de l'échantillon



+ Echantillon probabiliste



Echantillon aléatoire

Exemple : On interroge des passagers, qui descendent du train, au hasard dans les gares

Tous les individus de la population ont tous la même probabilité de faire partie de l'échantillon / d'être « tiré au sort »

Le choix est aléatoire

Indépendance des individus composant l'échantillon

D'après les lois de probabilités, l'échantillon est représentatif de la population



Il est nécessaire de disposer d'une liste de la population pour réaliser le tirage au sort



Echantillon proportionné à la taille

Exemple : On interroge aléatoirement les passagers dans les gares en enquêtant davantage les grandes gares

On pondère les individus de la population selon leur importance relative

Si une unité a plus de poids qu'une autre, sa probabilité d'être considérée augmente

Plus grande efficacité que les échantillons aléatoires



La définition de l'importance relative est cruciale



Echantillon systématique

Exemple : J'interroge la 1^{er} personne qui sort du train et tous les 20 passagers suivants

Chaque élément de l'échantillon est choisit de façon régulière dans la population.

Le 1^{er} élément est choisi de façon aléatoire.

L'intervalle est régulier.

Taille de l'échantillon prédéterminé



Le point de départ et la régularité de l'intervalle induisent un biais et l'échantillon n'est pas représentatif



Echantillon par degré

Exemple : Je choisis la ville, puis la gare puis le trajet puis le train et enfin un échantillon aléatoire des voyageurs de ce train

On réalise des tirages à différents niveaux.

Sélection de la 1^{ère} unité (unité primaire) puis, de façon aléatoire l'unité secondaire Jusqu'à l'identification de la dernière unité



Aucune représentativité



Echantillon par grappe

17

Exemple : La SNCF choisit aléatoirement 10 trajets de la journée et interroge tous les passagers de ces trains

Au défini des sous-zones dans une zone géographique définie.

Au sein de chaque sous-zone (grappe), on réalise un tirage aléatoire

Pas besoin de connaître les individus qui composent la population

Méthode appropriée lorsque la zone étudiée est large



Le découpage retenu peut induire un biais

Toutes les grappes n'ont pas le même nombre d'individus



Echantillon stratifié

Exemple : Compte tenu du trafic dans chaque gare, j'interroge un nombre de voyageurs différent d'une gare à l'autre (au prorata).

On découpe la population en strates

Sélection aléatoire dans chaque strate en respectant l'équation suivante :

$$\frac{\text{Individus enquêtés dans la strate}}{\text{Taille de l'échantillon}}$$

=

$$\frac{\text{Dimension totale de la strate}}{\text{Taille de la population}}$$

Echantillon représentatif de la population au regard des critères de stratification



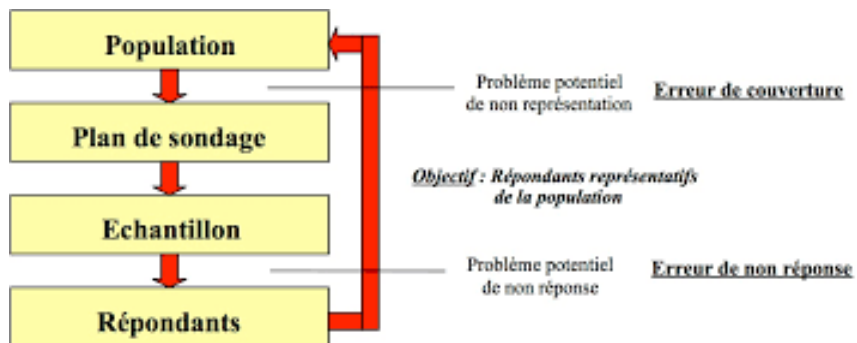
Besoin de connaître la structure de la population



+ Ce qu'il faut retenir



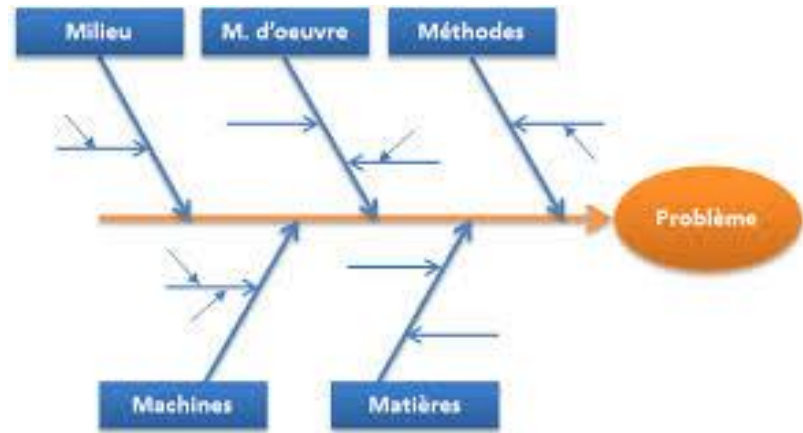
Ce qu'il vous fait retenir



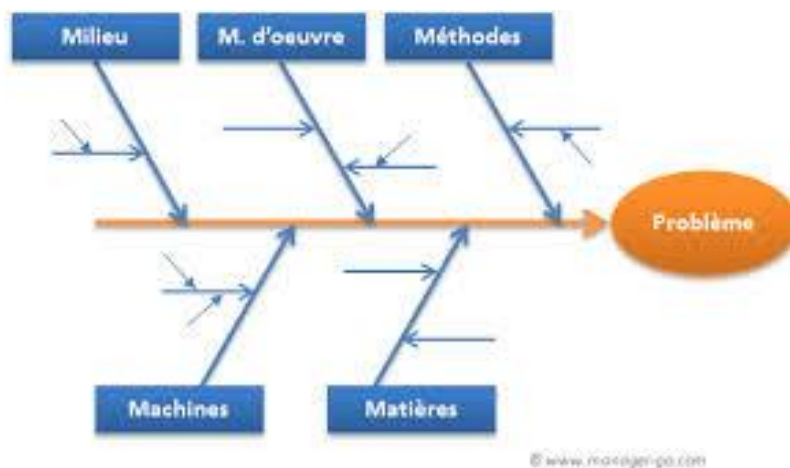
Dans tous les cas, vous devez être en mesure de justifier la méthode d'échantillonnage retenue



Recommandations pour éviter les biais : enquêteurs, enquêtés et environnement



© www.marcel-ga.com

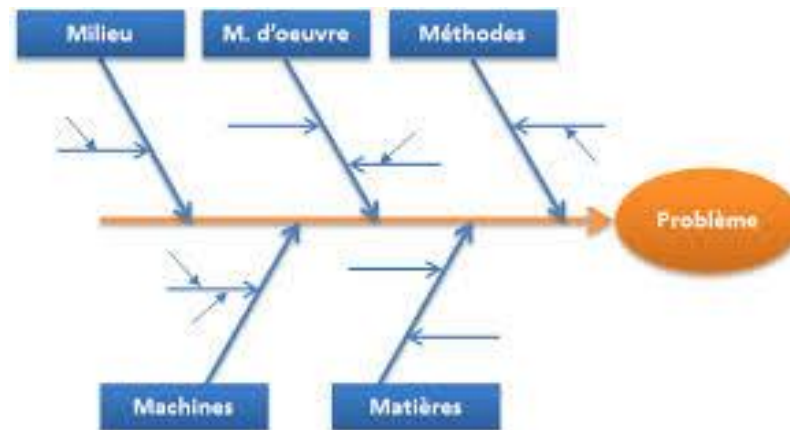


+ Anticipation des biais par la méthode des 5M

+ La méthode des 5M (1/2)

Créée par le professeur Kaoru Ishikawa (1915-1989) d'où son appellation « Méthode d'Ishikawa » ou 5M ou diagramme en « arêtes de poisson » ou diagramme de « causes à effet ».

Objectif est de rechercher et de représenter de manière synthétique les différentes **causes** possibles d'un problème (**effet**).





La méthode des 5M (2/2)

Une fois les causes identifiées, l'enjeu est de mettre en place des actions correctives appropriées.

C'est un outil correctif mais attention pas uniquement !

C'est un outil d'aide à l'identification préalable des éléments (causes) qui pourraient conduire à un biais qu'on veut éviter.



Anticipation des biais par la méthode des 5M

25

Un bon **questionnaire** (Support)



Matériel

Qui doit permettre de répondre à la
question de recherche



Méthode

Ce questionnaire est administré à des
enquêtés ...



Matière

Par des **enquêteurs** qui évoluent



Main d'œuvre

Dans un **environnement** spécifique



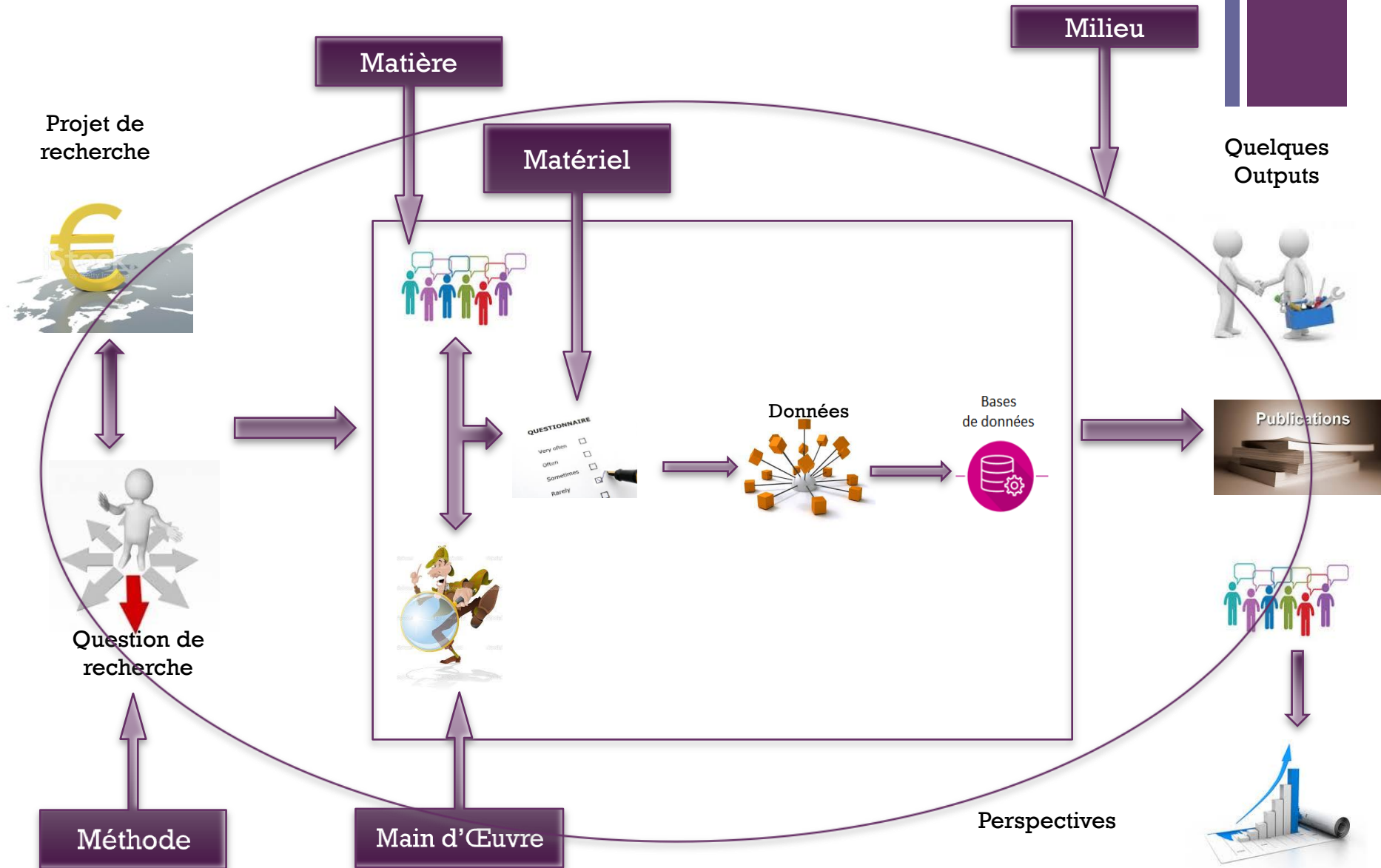
Milieu



La méthode des 5M

+ Articulation processus CVD & Méthode 5M

26





+ Biais liés au chercheur

Responsable du Traitement dans le cadre du RGPD

Acteur clef au cœur du dispositif d'enquête tout au long du cycle de vie des données



Biais liés au chercheur

Responsable du traitement dans le cadre du RGPD

- ❑ Les enquêteurs
 - ❑ Le nombre d'enquêteurs et de superviseurs dépend du nombre d'enquêtes à réaliser et du budget dont on dispose
 - ❑ Faut-il privilégier un recrutement en local ?
 - ❑ Faut-il privilégier la connaissance du terrain ou la compétence scientifique ?
 - ❑ Choisir des enquêteurs qui savent s'adapter & qui ont des qualités relationnelles
 - ❑ Recommandations claires aux enquêteurs
 - ❑ S'assurer de leur capacité à reformuler de façon neutre les questions (ne pas interpréter, ne pas orienter)
- ❑ Anticiper leur rémunération :
 - ❑ Ne pas rémunérer au nombre de questionnaires
 - ❑ Mise en œuvre, dans la mesure du possible, d'une prime à la qualité
 - ❑ Vérification en cours de saisie



Biais liés au chercheur

Responsable du traitement dans le cadre du RGPD

29

- ❑ Echantillonnage
 - ❑ Justification de la méthode d'échantillonnage en lien avec la question de recherche
 - ❑ Qui enquêter quand la personne identifiée est absente ?

- ❑ Phase test des questionnaires
 - ❑ La population enquêtée ne fait pas parti de l' échantillon final
 - ❑ Planification des enquêtes de ses enquêteurs

- ❑ Prise de contact avec les fournisseurs de données officielles
 - ❑ Importance des éléments de cadrage
 - ❑ Besoin d'informer les acteurs sur le terrain

Le M de Main d'œuvre



+ Biais liés à l'enquêteur / superviseur



Biais liés à l'enquêteur / superviseurs

La formation

- ❑ Formation initiale ou recommandations
- ❑ Bonne compréhension de la question de recherche et des enjeux sous-jacents
- ❑ Formation minimum de 4 jours avec des AR sur le terrain
- ❑ Capacité à reformuler de façon neutre les questions (ne pas interpréter, ne pas orienter)
- ❑ Apprendre aux enquêteurs à ne pas « forcer » l'enquêté à répondre, ni orienter ses réponses mais les guider
- ❑ Aptitude à respecter le déroulé du questionnaire
- ❑ Importance d'un contrôle intermédiaire avec 1 superviseur pour 5 enquêteurs en fonction du nombre de zones à couvrir et des distances à parcourir



Biais liés à l'enquêteur / superviseurs

Importance du duo savoir-être & savoir-faire

- Etablir un climat de confiance

- Savoir- être : importance de la communication non verbale
 - Vestimentaire
 - Approche des enquêtés
 - Langage ...

- Biais si plusieurs enquêteurs !

- Avoir une écoute « active »



Biais liés à l'enquêteur / superviseurs En début d'entretien

- ❑ Se présenter (soi, l'institut ...)
- ❑ Présenter la recherche et la finalité de la recherche
- ❑ Justifier le choix de la population
- ❑ Expliquer pourquoi l'enquêté a été sélectionné
- ❑ Insister sur la confidentialité et l'anonymat des données
- ❑ Indiquer la durée approximative du questionnaire



Biais liés à l'enquêteur / superviseurs

Le consentement

34

Demander dans tous les cas le consentement des personnes enquêtées (écrit, audio...). Nécessité de laisser vos coordonnées pour que les enquêtés puissent exercer notamment leur droit de retrait (Cf. RGPD).

Prévoir, dans la mesure du possible, une personne supplémentaire pour la prise de note pendant que l'enquêteur conduit l'entretien

Demander à l'enquêté s'il autorise l'enregistrement :

- S'il refuse : en prendre acte et ne pas enregistrer.
- S'il accepte : lancer l'enregistrement et reposer la question pour avoir une trace audio de cet accord.

Acceptez-vous que cet entretien soit enregistré ? [Enregistrer la réponse]

Merci d'avoir accepté que cet entretien soit enregistré.

*Nous souhaitons mentionner que cet enregistrement restera confidentiel
et ne sera pas diffusé.*

Les données resteront anonymes et nous respecterons leur confidentialité.

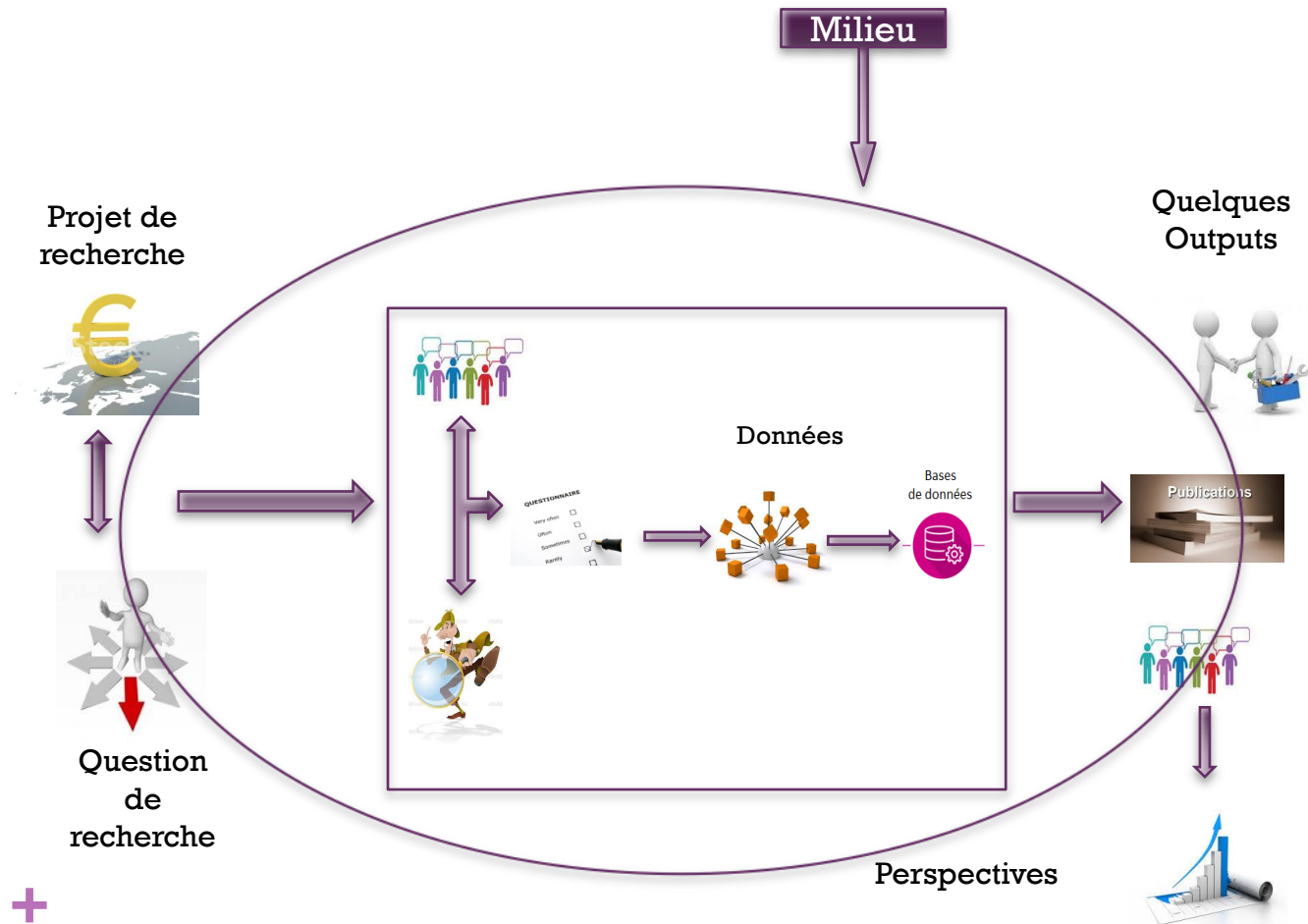


Biais liés à l'enquêteur / superviseurs

Quand l'évidence s'impose

- ❑ Remercier l'enquêté pour le temps accordé, son accueil ...
- ❑ Demander en fin d'enquête si les enquêtés acceptent d'être recontacté en cas de demande d'informations complémentaires
- ❑ Ne pas oublier que les enquêtés d'aujourd'hui sont les répondants de demain en cas de panel.
- ❑ Ne pas négliger l'effet « bouche à oreille » (ils échangent entre eux et les retours positifs (négatifs) des uns incitent (freinent) les autres à répondre au questionnaire

Le M de Milieu



Biais liés à l'environnement



Biais lié à l'environnement

S'adapter à un environnement différent du nôtre

- ❑ Adéquation entre la sensibilité de la question de recherche et l'environnement
- ❑ Formulation des questions et surtout leur reformulation en cas d'incompréhension
- ❑ Bonne adéquation entre nos outils de travail et l'environnement de l'enquête (stylo, feuille...)
- ❑ S'adapter / Respecter les us et coutumes du lieu de l'enquête



Biais lié à l'environnement

Quel que soit l'environnement

- ❑ Le temps local n'est pas le temps chercheur
- ❑ Interroger tous les enquêtés dans le même environnement
- ❑ Le moment le plus propice
- ❑ Ne pas avoir un questionnaire de « 30 pages » devant soi pour ne pas effrayer l'enquêté
- ❑ Eviter certains lieux : arrêt de tram...
- ❑ Ne pas demander à changer de lieux en cours d'enquête
- ❑ De façon générale faire attention aux éléments suivants : bruit, odeur, lieu...



+ Biais liés à l'enquête



Et l'enquêté ?

40

- ❑ Ce n'est ni le bon moment ni la « bonne » personne
 - ❑ Pas le bon moment donc réponse rapide
 - ❑ Identification de la population cible : si les personnes enquêtées ne correspondent à la population – envisager une façon polie de mettre un terme à l'entretien et apprendre à écourter le questionnaire
- ❑ Ne pas sous estimer l'effet « boule de neige »
 - ❑ Ne pas oublier que les enquêtés d'aujourd'hui sont les répondants de demain en cas de panel
 - ❑ Ne pas négliger l'effet « bouche à oreille » (ils échangent entre eux et les retours positifs (négatifs) des uns incitent (freinent) les autres à répondre au questionnaire

Le M de Matériel



+ Ne pas oublier que tout repose sur un bon questionnaire

Atelier du 23 mai
Construction d'un questionnaire



Ne pas oublier :

Les 5M pour limiter
les biais

Un bon **questionnaire** (Support)



Matériel

Qui doit permettre de répondre à la
question de recherche



Méthode

Ce questionnaire est administré à des
enquêtés ...



Matière

Par des **enquêteurs** qui évoluent



Main d'œuvre

Dans un **environnement** spécifique



Milieu



La méthode des 5M



Sensibiliser et former pour s'adapter aux enjeux actuels et pour améliorer collectivement nos pratiques

Pour se conformer aux exigences réglementaires du Règlement Général sur la Protection des Données (RGPD)

	Date & salle
Mise en œuvre du PDG et harmonisation/standardisation des métadonnées	Jeudi 11 avril 2019 – 09 h – 12 h Salle 1 – bât 15 - Cirad
2^{ème} édition Qu'est-ce que le RGPD ? Pourquoi sommes-nous concernés ?	Jeudi 18 avril – 14 h – 17 h Salle 2 – bât 15 - CIRAD

Les autres thématiques abordées en atelier

	Date & salle
Construction d'un questionnaire en lien avec la structuration de la base de données	Jeudi 23 mai - 14h à 17h Salle 1, Bât 15 – Cirad
Outils et méthodes d'auto-évaluation	Jeudi 06 juin- 14h à 17h Salle 1, Bât 15 - Cirad
Dépôt des données et métadonnées dans les entrepôts	Montpellier Occitanie